DOCUMENT RESUME

ED 445 065 TM 031 681

AUTHOR Baron, Patricia; Curley, Edward; Feigenbaum, Miriam
TITLE The Effects of Using Different Language Background Indicators on SAT I DIF Analyses.
PUB DATE 2000-04-00
NOTE 21p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 25-27, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Entrance Examinations; English; *High School Students; High Schools; *Item Bias; *Language Proficiency; Limited English Speaking
IDENTIFIERS *Scholastic Assessment Tests; *Student Descriptive Questionnaire

ABSTRACT

This study was conducted because revisions are being considered to the language questions that currently appear in the Student Descriptive Questionnaire (SDQ) of the Scholastic Assessment Test (SAT). The current test includes two questions about language acquisition: (1) "What language did you learn to speak first (EFL)?" and (2) "What language do you know best (EBL)?" Data were derived from a standard SAT I administration in 1998-1999 at which 192,737 high school juniors and seniors were tested. The study began by considering the effects of using samples derived from the EBL question rather than the current EFL-derived samples for differential item functioning (DIF) analyses. The first analysis was of Mantel Haenszel (MH) FID item statistics, which indicated that additional verbally-loaded mathematics and verbal questions would be flagged as inappropriate if EBL-derived samples were used. SDQ response patterns and scaled score data for several racial/ethnic groups were also examined to see if these data suggested any reasons for concern. Findings suggest that if the EFL question remains as it is now worded, then no change to current SAT I procedures would be necessary. The answers A (English only) and B (English and another language) to the EFL question would continue to define the target population for SAT I DIF analyses. If the EFL questions were dropped from the SDQ, then the analyses in this study would suggest, chiefly as a result of Asian American response patterns, that EBL A is better than EBL A and B to define the target population. However, for the Hispanic group, it would be undesirable to exclude the EBL B group from SAT I DIF analyses. It seems that the EBL question as it currently appears is inappropriate to define the target population for SAT I DIF analyses no matter which responses are used. Some suggestions are offered for additional research. (Contains 11 tables and 4 figures.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 065

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

P. Baron

1

THE EFFECTS OF USING DIFFERENT LANGUAGE BACKGROUND INDICATORS ON SAT I DIF ANALYSES

Patricia Baron

Edward Curley

Miriam Feigenbaum

Educational Testing Service

TM031681

Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA, April, 2000. The authors are listed alphabetically. Requests for reprints should be addressed to P. Baron, ETS, MS 15-L, Princeton, NJ 08541.

BEST COPY AVAILABLE

THE EFFECTS OF USING DIFFERENT LANGUAGE BACKGROUND INDICATORS ON SAT I DIF ANALYSES

The SAT I: Reasoning Test is designed to measure verbal and mathematical reasoning skills for college-bound high school juniors and seniors who have a reasonable level of proficiency with the English language. Part of the evaluation of the validity of a standardized test such as the SAT I includes the selection of appropriate samples of examinees on which to perform various statistical analyses. For the SAT I, most statistical analyses are based on the target sample of all juniors and seniors. For example, even though many examinees who take the SAT I are in grades 7, 8, 9, or 10, classical item analysis samples and equating samples are restricted to high school juniors and seniors tested under standard conditions.

The Student Descriptive Questionnaire (SDQ), which every examinee is asked to complete prior to taking the SAT I or the SAT II: Subject Tests, currently includes two questions about language acquisition: "What language did you learn to speak first?" (EFL) and "What language do you know best?" (EBL). There are three answer choices for both of the questions: "English only" (A), "English and another language" (B), and "Another language" (C). Student responses to these two language background questions, and to SDQ gender and ethnic/racial group membership questions, are used at Educational Testing Service (ETS) to define the groups of examinees on whom Differential Item Functioning (DIF) analyses are calculated for SAT I and SAT II.

DIF refers to a difference in item performance between two groups of examinees matched for ability with respect to the construct being measured by the test. DIF

analyses done at ETS allow test developers to evaluate the differential difficulty of items for various reference groups (White and male) and focal groups (African American, Asian American, Hispanic/Latino, Native American, and female). DIF analyses are intended to screen out items that may differentially advantage or disadvantage examinees based on their group membership rather than on their ability in the construct being measured. The target sample for DIF analyses for the SAT, therefore, must be examinees who have a level of proficiency in English that allows for the accurate assessment of their verbal and math reasoning skills. If, because of deficiencies in English, some examinees answer questions incorrectly that they have the knowledge to answer correctly, then those questions may show DIF inappropriately -- as a result of the examinees' level of English language proficiency rather than the examinees' group membership.

DIF analyses for the SAT I are conducted along with item analyses after newly written verbal and math questions are tried out ("pretested") in unscored, separately timed 30-minute sections during the three-hour testing session. Pretest questions that are too hard or easy, that discriminate poorly, or that show high levels of DIF are considered inappropriate for use in operational forms of the SAT I; such items are removed from SAT I pools, therefore, before operational forms are built. DIF analyses are also used after the administration of each new operational form to monitor the levels of DIF observed in the previously approved items and in the test as a whole.

SAT I has used the EFL question to define its target population since 1985, in anticipation of the introduction of DIF analyses to the program. (The EBL question had two response choices (Yes/No) prior to 1985, was dropped from the SDQ from 1985-89,

then was reintroduced in its present format in 1989 for use by SAT II.) Currently, SAT I includes in its DIF analyses all examinees who answer "English only" (A) or "English and another language" (B) in response to the EFL question. Those who answer "another language" (C) to the EFL question are excluded from SAT I DIF analyses.

Recently, ETS and the College Board (CB), sponsor of the SAT, have considered shortening the SDQ and possibly revising or replacing one or both of the language questions. Before such steps are taken, however, a thorough analysis of available data needs to be undertaken to try to inform whatever changes may be made. We have begun this research effort by doing some descriptive analyses of the EBL question. Our study is exploratory in nature and thus follow-up research will be needed. The data in this paper are derived from a standard SAT I administration at domestic test centers during the 1998-99 testing year, at which the following numbers of junior and senior examinees were tested.

Table 1: SAT I ETHNIC/RACIAL GROUP VOLUMES

White	133,083
African American	23,218
Asian American	18,395
Hispanic	18,041

Note: Native American volumes at this administration were too small to be included as part of this study.

To examine the effect of changing the sample on whom DIF analyses are performed, we have analyzed three kinds of data: item-level data, response patterns to the SDQ language questions, and scaled score data. When we looked at scaled score data, we considered mean scores for verbal (V) and for math (M) as well as differences

between mean scores (V-M) for each of the ethnic/racial groups in Table 1. Using the EFL and EBL questions to define different target samples of test takers allows for a comparison of the effect on DIF statistics of changing the sampling criteria.

At ETS, item difficulty estimates are computed in the delta metric, which has a mean of 13 and a standard deviation of 4. Holland and Thayer (1985) converted the Mantel Haenszel (MH) DIF statistic into a difference in the delta metric, referred to as MH D-DIF. Negative values of MH D-DIF mean that a question is differentially more difficult for the focal group; positive values mean that a question is differentially more difficult for the reference group. Initially for this study, MH D-DIF item statistics were calculated on two samples of examinees. The DIF statistics for those who answered either "English only" (A) or "English and another language" (B) to the EFL question (i.e. the sample on whom SAT I DIF analyses are currently run) were compared to the DIF statistics for those who responded either (A) or (B) to the EBL question, with the following results.

Table 2:
CORRELATIONS OF EFL-DERIVED MH D-DIF VS. EBL-DERIVED MH D-DIF

	Male/Female	White/African American	White/Asian American	White/Hispanic
Verbal	0.996	0.993	0.924	0.911
Math	0.990	0.987	0.938	0.891

From these correlations, it is evident that Hispanic and Asian American examinees responded more differently to the EFL and the EBL questions than did African American or female examinees.

ETS uses the MH D-DIF statistic to classify SAT I pretest items into one of three categories: those showing negligible, slight to moderate, or moderate to large levels of differential item functioning (Zieky, M., 1993). Items in the third category are removed as unacceptable from the operational pools of verbal and math items. Based on the above correlations we looked, for Hispanic and Asian American examinees, at those particular pretest items that were classified into different DIF categories when we used EFL A+B responses versus EBL A+B responses to define the target population of our DIF analyses. Overall from the studied SAT I administration, 30 of 224 verbal pretest items and 16 of 200 math pretest items shifted DIF categories for one or both of these two groups of examinees (about 11% of the total pretest items).

We found that none of the items that were actually to be removed from the pools with unacceptable levels of MH D-DIF using EFL A+B responses for Hispanic or Asian American examinees would have remained in the pools using EBL A+B responses to define the target populations. Instead, we found that using the EBL question to screen for DIF resulted in an increased number of items showing moderate to large amounts of DIF. Those items classified with unacceptable levels of MH D-DIF using EBL A+B responses but not classified as such using EFL A+B responses included the following: for SAT-Verbal, a few easy and hard Analogies and Sentence Completions and several Reading items measuring vocabulary in context; for SAT-Math, several verbally-loaded word problems. Why were these sorts of items being flagged as inappropriate for operational pools when using the EBL question but not when using the EFL question to define the target populations for Hispanic and Asian American examinees?

A look at the numbers of examinees by ethnic/racial group who answered "English only" (A) or "English and another language" (B) to the EFL and EBL questions revealed some interesting results. (Remember that every SAT I examinee is asked to answer both of these questions.)

Table 3:
EXAMINEE RESPONSES TO EFL and EBL QUESTIONS BY ETHNIC/RACIAL GROUP

White (133,083)		**Asian American (18,395)**	
EFL A or B	130,397 (98%)	EFL A or B	11,197 (61%)
EBL A or B	132,331 (99%)	EBL A or B	16,403 (89%)
EBL A (only)	129,567 (97%)	EBL A (only)	12,181 (66%)
African American (23,218)		**Hispanic (18,041)**	
EFL A or B	22,562 (97%)	EFL A or B	12,312 (68%)
EBL A or B	23,005 (99%)	EBL A or B	17,139 (95%)
EBL A (only)	21,995 (95%)	EBL A (only)	11,219 (62%)

Table 3 reveals that, for the White and African American groups, there is very little difference between the combined numbers of examinees who answer (A) or (B) to the EFL question and those who answer (A) or (B) to the EBL question. For these groups, virtually everyone answers "English" or "English and another language" to both the first and the best language questions. However, the response patterns for the Asian American and Hispanic groups are markedly different. For these two groups, the numbers of examinees who answer (A) or (B) to the EFL question are much more similar to the numbers who answer (A) to the EBL question than they are to the numbers who answer (A) or (B) to the EBL question. Said another way, many fewer

Asian American and Hispanic examinees answer "Another language" (C) to the English best language question than they do to the English first language question.

We can only hypothesize about the reason(s) for such response patterns. Certainly the EBL question is more subjective than the EFL question. Perhaps many Asian American and Hispanic examinees have developed their English skills so much over the years that they perceive themselves to be bilingual when in fact their language proficiencies still differ. Or perhaps the question "What language do you know best?" seems to be a high-stakes question to examinees seeking admission to English-speaking colleges and universities, and thus some answer it the way they believe colleges would want them to answer. (In fact, examinees are told that some of their individual SDQ responses will be shared with colleges.) In any case, it seems relevant to look closely at response patterns to the SDQ questions (particularly for these two groups of examinees) to inform decisions about possible changes to the language questions used to determine the SAT I DIF target population.

Tables 4 through 11 provide the numbers of examinees and their mean scaled scores, for verbal and for math, for those Asian American, Hispanic, African American, and White examinees who selected each of the three responses to the two different language questions on the SAT SDQ. (The SAT scale for verbal and for math runs from a low of 200 to a high of 800, with a mean near 500 and a standard deviation of about 110. Standard deviations associated with the mean scores reported in Tables 4 and 5 ranged from about 92 to 123; in Tables 6 and 7, from about 72 to 100; in Tables 8 and 9, from about 75 to 101; and in Tables 10 and 11, from about 80 to 112.)

Table 4:
SAT-VERBAL SCALED SCORE SUMMARY STATISTICS
FOR ASIAN AMERICAN EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ↓
English First (A)	N	4,996	140	50	5,186
	MEAN	521	451	414	519
English & Another First (B)	N	4,202	1,635	173	6,011
	MEAN	530	459	352	506
Another First (C)	N	2,983	2,447	1,768	7,198
	MEAN	517	440	366	454
EBL → Sums	N	12,181	4,222	1,991	18,395
	MEAN	524	448	366	489

Table 5:
SAT-MATH SCALED SCORE SUMMARY STATISTICS
FOR ASIAN AMERICAN EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ↓
English First (A)	N	4,996	140	50	5,186
	MEAN	543	478	509	541
English & Another First (B)	N	4,202	1,635	173	6,011
	MEAN	567	510	515	550
Another First (C)	N	2,983	2,447	1,768	7,198
	MEAN	562	524	561	549
EBL → Sums	N	12,181	4,222	1,991	18,395
	MEAN	556	517	556	547

BEST COPY AVAILABLE

Table 6:
SAT-VERBAL SCALED SCORE SUMMARY STATISTICS
FOR HISPANIC EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ↓
English First (A)	N	5,774	313	29	6,116
	MEAN	491	440	364	488
English & Another First (B)	N	3,719	2,405	72	6,196
	MEAN	469	439	377	456
Another First (C)	N	1,726	3,202	800	5,729
	MEAN	467	432	385	436
EBL → Sums	N	11,219	5,920	901	18,041
	MEAN	480	435	384	461

Table 7:
SAT-MATH SCALED SCORE SUMMARY STATISTICS
FOR HISPANIC EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ↓
English First (A)	N	5,774	313	29	6,116
	MEAN	486	424	373	482
English & Another First (B)	N	3,719	2,405	72	6,196
	MEAN	464	433	396	451
Another First (C)	N	1,726	3,202	800	5,729
	MEAN	467	440	429	446
EBL → Sums	N	11,219	5,920	901	18,041
	MEAN	476	436	424	460

Table 8:
SAT-VERBAL SCALED SCORE SUMMARY STATISTICS
FOR AFRICAN AMERICAN EXAMINEES

		English Best (A) ◄	English & Another Best (B)	Another Best (C)	EFL Sums ▼
English First (A)	N	20,335	389	60	20,784
	MEAN	448	424	388	447
English & Another First (B)	N	1,385	369	24	1,778
	MEAN	435	414	313	429
Another First (C)	N	275	252	129	656
	MEAN	443	408	341	410
EBL → Sums	N	21,995	1,010	213	23,218
	MEAN	447	416	351	445

Table 9:
SAT-MATH SCALED SCORE SUMMARY STATISTICS
FOR AFRICAN AMERICAN EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ▼
English First (A)	N	20,335	389	60	20,784
	MEAN	433	402	383	432
English & Another First (B)	N	1,385	369	24	1,778
	MEAN	424	417	362	421
Another First (C)	N	275	252	129	656
	MEAN	434	423	401	423
EBL → Sums	N	21,995	1,010	213	23,218
	MEAN	433	413	391	431

BEST COPY AVAILABLE

Table 10:
SAT-VERBAL SCALED SCORE SUMMARY STATISTICS
FOR WHITE EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ↓
English First (A)	N	124,595	940	141	125,676
	MEAN	523	504	457	523
English & Another First (B)	N	4,016	669	36	4,721
	MEAN	495	478	415	492
Another First (C)	N	956	1,155	575	2,686
	MEAN	510	484	425	481
EBL → Sums	N	129,567	2,764	752	133,083
	MEAN	522	490	431	521

Table 11:
SAT-MATH SCALED SCORE SUMMARY STATISTICS
FOR WHITE EXAMINEES

		English Best (A)	English & Another Best (B)	Another Best (C)	EFL Sums ↓
English First (A)	N	124,595	940	141	125,676
	MEAN	522	491	469	521
English & Another First (B)	N	4,016	669	36	4,721
	MEAN	498	482	479	496
Another First (C)	N	956	1,155	575	2,686
	MEAN	527	532	536	531
EBL → Sums	N	129,567	2,764	752	133,083
	MEAN	521	506	521	521

To read these tables most effectively, start in the lower right corner. For example, in Tables 4 and 5, the total number of Asian American examinees is 18,395 (all of the junior and senior Asian American examinees tested at the studied SAT I administration). Their mean verbal score was 489 and their mean math score was 547.

Looking at the number of Asian American examinees in the far right column of Tables 4 and 5, note that 5,186 answered "English first" (EFL A) and 6,011 answered "English and another first" (EFL B). These were the 11,197 Asian American examinees on whom operational SAT I DIF analyses were actually run for both the verbal and math tests. Note that the mean verbal scores for these two groups of examinees are relatively close (519 vs. 506), a 13-point difference. Similarly, their mean math scores are close (541 vs. 550), only a 9-point difference between those who answered "English first" and those who answered "English and another first."

Next, look across the bottom row of each table; in Tables 4 and 5, note that 12,181 Asian American examinees answered "English best" (EBL A) but only 4,222 answered "English and another best" (EBL B). Note also that the mean verbal scores for these two groups were quite discrepant (524 vs. 448), a 76-point difference. Their mean math scores are also discrepant (556 vs. 517), a 39-point difference between those who answered "English best" and those who answered "English and another best."

Tables 6 through 11 can be read in similar fashion, but it is the information on Asian Americans -- and particularly the differences between mean verbal and math scores for the groups of Asian American examinees who chose the various responses to the EFL and EBL questions -- that is of greatest significance to the issues addressed

in this paper. There is clearly an English-language component to SAT-Verbal; the assessment of verbal reasoning in English requires proficiency in English. There is also an English-language component to SAT-Math, though obviously it is smaller than that of SAT-Verbal. If the Asian American examinees who responded "English and another best" had mean SAT scores similar to those of the Asian American examinees who responded "English best" (or similar to the groups who responded "English first" or "English and another first" -- the current sample on whom DIF analyses are run), then it might seem that the examinees who responded "English and another best" have a level of English proficiency that warrants including them in the SAT I DIF analyses. However, this is not the case. In fact, the mean verbal score of those Asian Americans who answer "English and another best" (448) is lower than the mean verbal score of those who answer that they learned "Another (language) first" (454). Furthermore, the mean math score of those Asian Americans who answer "English and another best" (517) is actually lower than the mean math score of <u>any</u> of the other groups of Asian American respondents to either of the two language questions. Is there a deficiency of English language skills for those who respond "English and another best" that may be affecting even their math scores?

Figures 1 through 4 present, separately for each of the four ethnic/racial groups, <u>differences</u> between verbal and math mean scaled scores for selected combinations of responses to the EFL and EBL questions. **[Insert figures here]** Figure 1 combines the identical verbal and math data for Asian American examinees found in Tables 4 and 5, Figure 2 does the same for Hispanic examinees in Tables 6 and 7, etc. Note in the figures that, if the bars are <u>below</u> the line of zero differences, then the mean <u>math</u> score

is higher than the mean verbal score for that group of respondents; if the bars are above zero, then verbal scores are higher than math. In each figure, the three bars on the left represent V-M mean scaled scores for those EBL A examinees who also answered (respectively) EFL A, EFL B, or EFL C. Similarly on the right side of each figure for the EBL B examinees.

In an attempt to keep these figures as simple as possible, verbal and math scaled score differences for examinees who answered EBL C are not included for two reasons. First, examinees who respond that they know another language better than they know English would not be included in the SAT target population for DIF analyses in any case. Second, as can be seen in Tables 4 through 11, the number of EBL C respondents who also answer EFL A or EFL B is only 1% or less of each of the four ethnic/racial groups. (As might be expected, mean math scores in most cases are much higher than mean verbal scores for any group of examinees who responds EBL C.)

We have indicated in each figure the percentage of the total ethnic/racial group who answered both EBL B and EFL C. Note that for African American and White examinees this percentage is very small (1%), but for Asian Americans (13%) and Hispanics (18%) the proportion of examinees who answer both EBL B and EFL C is much larger. Note also that, for Hispanic examinees who answer both EBL B and EFL C, the actual difference between mean verbal and mean math scores is quite small, only 8 scaled score points (as indicated in Tables 6 and 7: 3,202 examinees, verbal=432, math=440). On the other hand, Asian Americans who answer both EBL B

and EFL C have mean verbal scores that are 84 points lower than their mean math scores (as indicated in Tables 4 and 5: 2,447 examinees, verbal=440, math-524).

If the EBL A group and the EBL B group are both proficient in English, then we should expect the differences between their verbal and math scores to be similar. On the other hand, if the EBL B group is actually less proficient in English than the EBL A group, then there should be a greater effect on their verbal scores than on their math scores, which is exactly what we see for Asian Americans in Figure 1. (Note that, although the 2,983 Asian American examinees who responded EBL A and EFL C have a 45-point difference between their mean verbal and math scores (verbal=517, math=562), their mean verbal score of 517 is very similar to the mean verbal score of 521 for the 4,996 Asian American examinees who answer both EBL A and EFL A.)

In conclusion, it appears that the EBL B group, especially for Asian American examinees, includes a number of students for whom English is not really one of their best languages. The effect of including students in the SAT I DIF analyses who are not reasonably proficient in English is that items (such as vocabulary in context and math word problems) will be flagged for DIF because they appear to be disadvantaging an ethnic/racial group based on group-specific rather than construct-specific factors. However, since this group actually lacks proficiency in English, it is instead likely that items will be flagged due to English-language factors. If the response pattern to the EBL question from this studied administration is representative then, over the years, dozens or even hundreds of math and verbal questions could be deleted from SAT I verbal and math pools inappropriately.

SUMMARY

This research was conducted because revisions are being considered to the language questions that currently appear in the SAT SDQ. We began by exploring the effects of using EBL-derived samples rather than the current EFL-derived samples for SAT I DIF analyses. We looked first at MH D-DIF item statistics, which indicated that additional verbally-loaded math and verbal questions would be flagged as inappropriate if we used EBL-derived samples. We next looked at SDQ response patterns and scaled score data for several ethnic/racial groups to see if these data suggested any reasons for concern. In the end, we believe that we have answered some important questions (and raised some important new ones) about the effects of using different language background indicators on SAT I DIF analyses.

If the EFL question were to remain as it is now worded, then no change to current SAT I procedures would be necessary: we would continue to use EFL A+B as the target population for SAT I DIF analyses. If the EFL question were dropped from the SDQ, then the analyses in this paper would seem to suggest -- due chiefly to Asian American response patterns -- that EBL A is better than EBL A+B to define the target population. However, for the Hispanic group it would, in fact, be undesirable to exclude the EBL B group from SAT I DIF analyses. Note in Tables 6 and 7 that 5,920 Hispanic examinees answer "English and another best" (EBL B), almost one-third of the total Hispanic group. Note also that the mean verbal score (435) and the mean math score (436) of these 5,920 examinees are almost identical (V-M=1), similar to the difference between the verbal mean (480) and the math mean (476) of the EBL A Hispanic

respondents (V-M=4). This similarity suggests that, for Hispanic examinees who answer EBL B, English is indeed one of their best languages.

It seems then that the EBL question as it currently appears in the SAT SDQ is inappropriate to define the target population for SAT I DIF analyses no matter which response(s) are used. Several follow-up steps to this initial exploratory analysis are necessary, however, before drawing any firm conclusions. We recommend first that statistical tests be performed to evaluate the significance of the differences reported in this paper among various DIF data, mean scaled scores, and SDQ response patterns. Second, we recommend that our findings about the EBL question be substantiated by looking at data from additional SAT I administrations. Finally, if these two steps suggest that our initial findings are indeed valid and generalizable, then we recommend that research be conducted on ethnic/racial response patterns to alternative language background indicators. (See the Appendix for some proposed alternative language questions that might be used to collect data in the future from SAT I examinees.) Until all of the above research is concluded, we recommend that EFL A+B continue to be used to define the SAT I target population for performing DIF analyses.

SAT I VERBAL SCALED SCORE (SS) MEAN - MATH SCALED SCORE (SS) MEAN BY
RESPONSE TO ENGLISH FIRST LANGUAGE (EFL) AND ENGLISH BEST LANGUAGE (EBL) QUESTIONS

Figure 1:
ASIAN AMERICAN EXAMINEES

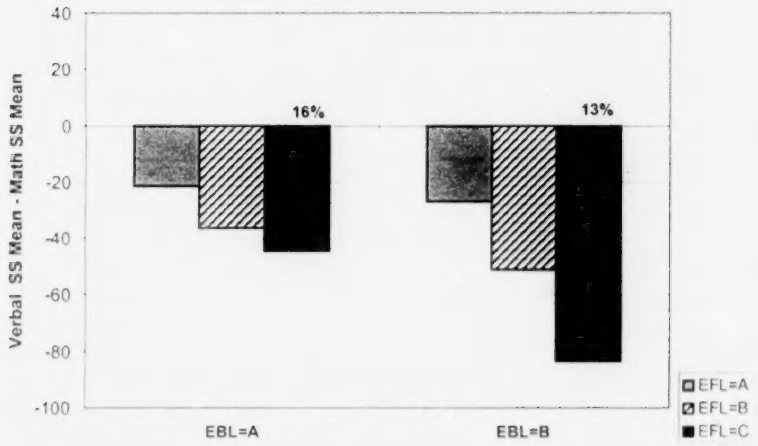


Figure 2:
HISPANIC EXAMINEES

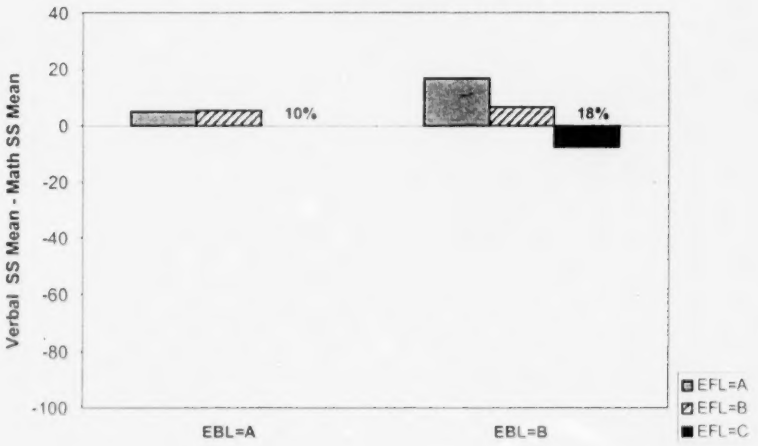


Figure 3:
AFRICAN AMERICAN EXAMINEES

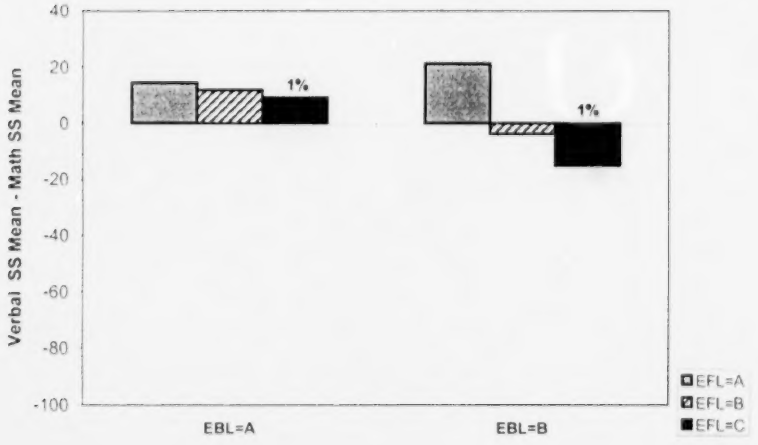
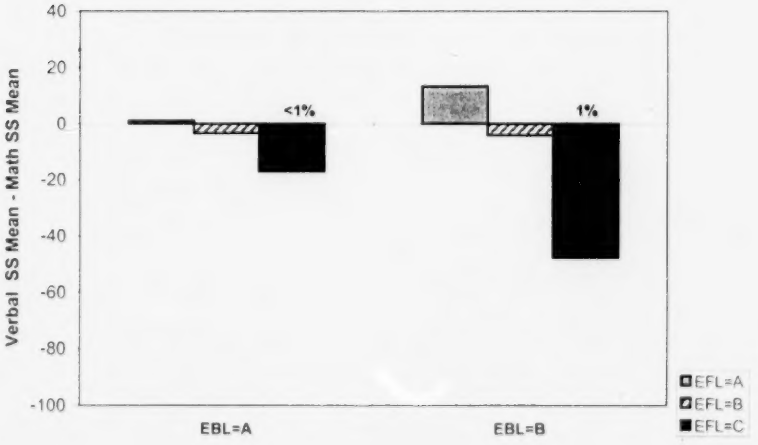


Figure 4:
WHITE EXAMINEES



BEST COPY AVAILABLE